Listeners integrate semantic and phonetic cues during sentence comprehension

Scarlet Wan Yee Li[1], Margarethe McDonald[2] and Tania Zamuner[1]

[1]University of Ottawa (Canada), [2]University of Kansas (United States)

Listeners actively interpret incoming information by coordinating different levels of linguistics cues when processing sentences within a discourse. Researchers have found that adult listeners can use either low-level phonetic cues (e.g., coarticulation) [3, 4, 5] or higher-level cues (e.g., morphosyntactic and semantic information) [1, 8] to interpret the unfolding speech signal and predict future parts of a sentence. While these studies have focused more on the processing of single cues during spoken word recognition, recent studies have also examined how multiple linguistic cues are combined and integrated during speech processing. Martin (2016) proposed a cue integration framework suggesting that listeners can flexibly use and weigh available cues across all levels of linguistic representations during language comprehension [7]. Behavioral studies investigating Martin (2016)'s cue integration framework showed that listeners can dynamically combine and integrate some of the lower-level and higher-level cues during spoken language comprehension [2, 6]. However, it is still not well understood how the mechanism of cue combination and integration works across different levels of linguistic representations.

The current pre-registered study examined how adults (n = 52) process a previously unstudied combination of cues – preceding higher-level semantic cues (i.e., semantic context) and later low-level acoustic cues (i.e., coarticulation) – during online spoken comprehension within Martin (2016)'s cue integration framework using an eye-tracking paradigm. Participants were tested on sentences that contained a prime (semantically related or semantically unrelated to the target) and a target which had varying coarticulation cues (matching vs. mismatching splicing cues). For example, *The man sees the river*$_{\textit{Semantically related}}$ */cartoon*$_{\textit{Semantically unrelated}}$ *(Prime) and looks for the boa*$_{t}$*t*$_{\textit{Matching splicing}}$ */ boa*$_{n}$*t*$_{\textit{Mismatching splicing}}$ *(Target).* Participants were presented with two pictures (e.g., target – boat and competitor – bone) on a screen. Analyses looked at the proportion of looking to the target during different time periods: prime and target windows.

Results based on the pre-registered analyses demonstrate that adults flexibly use both the preceding semantic cues and later coarticulatory cues once the cues are available. In the prime window, adults anticipatorily fixated to targets significantly more when they heard a semantically related prime compared to a semantically unrelated word ($p < .001$) (Figure 1A). In the target window, adults showed significantly greater proportional looking to targets containing matching splicing cues than targets containing the mismatching splicing cues in both semantic-related and semantic-unrelated conditions ($p < .001$) (Figure 1B). In an exploratory analysis that extended the target window timeframe, an interaction between semantic and splicing condition was found ($p = .033$) (Figure 1B). This analysis indicates that adults flexibly weighed both the preceding higher-level and later lower-level cues, such that the processing of low-level coarticulatory cue varied depending on the semantic context. To be specific, adults showed less sensitivity to the mismatching coarticulation cues (e.g., *boa*$_{n}$*t*) when the semantic context was semantically related to the target (e.g., *boat*). Adding to the previous findings within the Martin (2016) framework, our results provide evidence showing that the time course of the interaction between cues across levels of linguistic representations may differ; moreover, we provide insight on the set of cues that our cognitive system can combine and integrate during language comprehension by examining the combination of an unstudied level of cues.

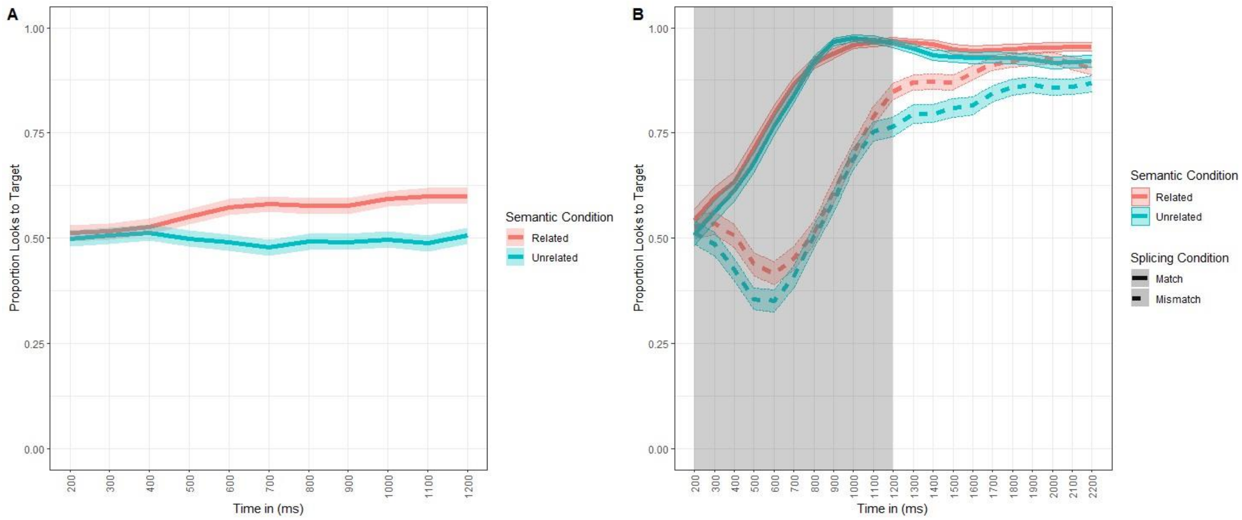

Fig. 1. Time course of proportion of looks to the target from 200 ms to 1200 ms after prime word onset (A: Prime Window). Time course of proportion of looks to the target from 200 ms to 2200 ms after target word onset. Pre-registered analyses shaded in grey (B: Target Window).

References

[1] Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of sentence reference. *Cognition*, *73*(3), 247– 264. https://doi.org/10.1016/S0010-0277(99)00059-1

[2] Bai, F., Meyer, A. S., & Martin, A. E. (2022). Neural dynamics differentially encode phrases and sentences during spoken language comprehension. *PLoS Biology*, *20*(7), e3001713. https://doi.org/10.1371/journal.pbio.3001713

[3] Beddor, P. S., McGowan, K., Boland, J., Coetzee, A., & Brasher, A. (2013). The timecourse of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*(4), 2350–2366. https://doi.org/10.1121/1.4794366

[4] Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534. https://doi.org/10.1080/01690960143000074

[5] Desmeules-Trudel, F., Zamuner, T.S. (2019) Gradient and categorical patterns of spoken word recognition and processing of phonetic details. *Attention, Perception, and Psychophysics*, *81*(5), 1654–72. https://doi.org/10.3758/s13414-019-01693-9

[6] Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020a). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(3), 549–562. https://doi.org/10.1037/xlm0000744

[7] Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*. *7*, 120. http://doi.org/10.3389/fpsyg.2016.00120

[8] Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1-14. https://doi.org/10.1037/0278-7393.32.1.1